

DSSTox Log File:

EPA Water Disinfection By-Products with Carcinogenicity Estimates (DBPCAN)

(last updated 10 April 2006)

Description: Information in this file documents creation, review, and update process for the DBPCAN SDF file, provides summary information on database content, and lists currently unavailable CAS registry numbers for known structures. The first section summarizes the process used for creating the initial DSSTox SDF files and the quality assurance checks and procedures employed. A table providing field and data counts offers summary overview of DBPCAN file content. The Log Table documents modifications and revisions to the database content or format in version updates. To obtain the most current version of this Log File and a record of any new modifications, or to report errors in this file, a user should consult the DSSTox DBPCAN database page: http://www.epa.gov/nheerl/dsstox/sdf_dbpcan.html

QA and Development Notes for v1a:

Data included in DBPCAN underwent a series of quality review checks prior to publication of initial launch version. The original database was obtained from reference materials associated with the Main Citation and from communication with the Source and main authors of that study. We thank Yin-tak Woo for providing the most current data tables in electronic form. Chemical structures were initially obtained by conversion of SMILES codes provided in the original database using CambridgeSoft's ChemOffice 2002 ChemDraw (ver 7.0 for Windows) for MS Excel. Structures were further verified against drawn structures and names in Source-provided materials. The ChemFinder website (<http://chemfinder.cambridgesoft.com/>) was used for checking CAS-to-structures and for retrieving CAS numbers for analogs of the predicted chemical. CambridgeSoft ChemFinder (ver 7.0 for Windows) was also used for automatic generation of SMILES codes from structures and both ChemFinder and ACD ChemFolder (ver 6.0 for Windows) were employed for "Structure-to-Name" or "Name-to-Structure" features. Original SMILES codes provided by the Source were converted to ChemFinder-generated SMILES when the former could not be converted to a Structure and verified within the ChemFinder application.

Notes for v2a:

For version 2a, a variety of fields have been added. IUPAC systematic chemical names, **ChemName_IUPAC**, were computed by Marc Nicklaus (NCI) using the ACD Labs IUPAC Name-Generation software (ACD/NameBatch, version 8.05). **INChI** codes were automatically generated from the final DSSTox SDF using a pre-release version of the publicly available program, wINChI11b.exe, accessible from the NIST INChI developers (<http://chemdata.nist.gov/IChI/INChIv11b.zip>). AuxInfo strings, which can be used to reproduce the molfile structure, are typically generated along with the INChI codes. However, due to their length frequently exceeding the 255 character limit of some Chemical Relational Database applications and the non-unique nature of the AuxInfo text string, we include only the invariant INChI codes in the DSSTox data files

Notes for v3a,b:

Revised DSSTox Standard Chemical Fields are included (see <http://www.epa.gov/nheerl/dsstox/MoreonStandardChemFields.html>) along with updated InChI codes (version 1.0), recomputed IUPAC chemical names (ACDLabs ACD/Name, version 9.0), and many regenerated 2D structures with stereochemistry of steroidal compounds rendered in more standardized form. Additionally, an extensive quality review of all DSSTox chemical records was performed, resulting in numerous corrections and modifications to chemical structures and added information (CASRN, representative structures for mixtures, etc) throughout DSSTox data files. For more information on current review procedures, see <http://www.epa.gov/nheerl/dsstox/ChemicalInfQAProcedures.html>

Log of SDF Modifications and Version/revision updates:

Date	DSSTox SDF File Name	Modifications from previous version	Additional Notes
12Sep03	DBPCAN_v1a_209_12Sep03	Initial launch publication; no previous published versions.	DBPCAN is considered a "static" historical database meaning that further expansion of the database to include additional data is unlikely. Future updates will correct reported errors provided by users or incorporate DSSTox format changes.
1Mar05	DBPCAN_v2a_209_1Mar05	New Fields: INChI , ChemName_IUPAC , StudyType , Endpoint Modified Field Names: DBPNote to ToxNote ; ChemName_Synonym to ChemName_Other ; CAS for DSSTox_ID=78,79 corrected; CAS added for 8 structures, DSSTox_ID=16, 25, 52, 74, 111, 125, 195, 200; SMILES and structure corrected for DSSTox_ID=25, 75, 121,122 (latter 2 switched) CAS incorrect for DSSTox_ID=24, changed to NOCAS	Major format modification to include INChI, IUPAC names, and ToxML fields.
10Apr2006	DBPCAN_v3b_209_10Apr2006	Updated with new DSSTox Standard Chemical Fields and entries (<i>revised Aug 2005</i>). Updated InChI codes (version 1.0). Updated IUPAC chemical names (ACDLabs Name to Structure, version 8.0). Expanded "ddmmmyear" format for dates in DSSTox file names (e.g., 10Apr2006). Revised DBPCAN Source-Specific Fields: ChemClass DBP changed to ChemClass_DBP ConcernLevel changed to ActivityConcernLevel_Carcinogenicity ; Rationale changed to ActivityConcernLevel_Rationale ; Rationale changed to ActivityConcernLevel_RationaleSource ; AnalogChemName to Analog_ChemicalName ; AnalogCAS to Analog_CASRN ; AnalogSMILES to Analog_SMILES ToxNote to ToxicityNote with expanded "NOD" entry.	Numerous structure modifications and changes in stereochemical rendering throughout DSSTox data files following major quality review. DBPCAN_v3a_209_22Oct2005: Note: earlier version of this file was provided to PubChem, with identical format to v3b but latter has undergone additional QA review and has a few minor structural corrections/modifications.

Field and Data Counts in DSSTox SDF file: Refer to DBPCAN_FieldDefFile for definitions and explanations of all terms.

DSSTox SDF	Standard Chemical Fields	Standard ToxML Fields	Source-specific fields	Chemical records total	STRUCTURE_ChemicalType:		
					Defined Organic	Inorganic	Parent
DBPCAN_v1a	11	0	9	209*			
DBPCAN_v2a	14	2	8	209*			

*Three pairs of replicate 2D structures exist in the database (cis/trans isomer pairs).

DBPCAN SDF Content*	Totals_v3b
# Records	209
DSSTox Standard Chemical Fields	18
DSSTox Standard Toxicity Fields	2
DBPCAN Source Fields	8
Total # Fields	28
Chemical Content	Counts_v3b
STRUCTURE_ChemicalType:	
defined organic	204
inorganic	5
organometallic	0
no structure	0
STRUCTURE_TestForm_ DefinedOrganic:	
parent	204
complex	0
salt	0
salt complex	0

TestSubstance_Description:	
single chemical compound	208
defined mixture or formulation	1
undefined mixture	0
macromolecule	0
unspecified or multiple forms	0

Wanted!! CASRN Information

The **Unknown** CASRN entries below are primarily an indication of the unstudied nature of many of the disinfection by-product entries in the DBPCAN database. However, if a user has information pertaining to any **Unknown** CASRN in the below listing, please report this using a [DSSTox Error Report Form](#) that can be accessed from any DSSTox SDF Download Page, and be sure to indicate all relevant information (full **DSSTox_ID_FileName**, **TestSubstance_ChemicalName**, nature of missing information, source of correct information, etc.). Thank you!

DSSTox_ID_FileName	TestSubstance_ChemicalName	STRUCTURE_SMILES	CASRN	Date of Request
13 DBPCAN_v3b_209_10Apr2006	1,2-Bis(1-methylethenyl)-benzene	<chem>C1=C(C(C)=C)C(C(C)=C)=CC=C1</chem>	Unknown	10Apr2006
23 DBPCAN_v3b_209_10Apr2006	Bromochloromethyl acetate	<chem>BrC(OC(C)=O)Cl</chem>	Unknown	10Apr2006
24 DBPCAN_v3b_209_10Apr2006	1,1-Bromochloropropanone	<chem>CC(C(Cl)Br)=O</chem>	Unknown	10Apr2006
29 DBPCAN_v3b_209_10Apr2006	1-Bromo-1,1-dichloropropanone	<chem>O=C(C)C(Cl)(Br)Cl</chem>	Unknown	10Apr2006
31 DBPCAN_v3b_209_10Apr2006	3-Bromopropylchloromethyl ether	<chem>BrCCCOCCl</chem>	Unknown	10Apr2006
40 DBPCAN_v3b_209_10Apr2006	tert-Butyl maleic acid	<chem>O=C(O)C=C(C(C)(C)(C))C(O)=O</chem>	Unknown	10Apr2006
44 DBPCAN_v3b_209_10Apr2006	3-Chloro-4-(bromochloromethyl)-5-hydroxy-2(5H)-furanone	<chem>O=C1OC(O)C(C(Br)Cl)=C1Cl</chem>	Unknown	10Apr2006
47 DBPCAN_v3b_209_10Apr2006	3-Chloro-2-butanol acetate	<chem>CC(C(OC(C)=O)C)Cl</chem>	Unknown	10Apr2006
48 DBPCAN_v3b_209_10Apr2006	2-Chlorobutenedioic acid	<chem>OC(C=C(C(O)=O)Cl)=O</chem>	Unknown	10Apr2006
50 DBPCAN_v3b_209_10Apr2006	Chlorodibromoacetaldehyde	<chem>BrC(Br)(C=O)Cl</chem>	Unknown	10Apr2006
53 DBPCAN_v3b_209_10Apr2006	2-Chloro-3-(dichloromethyl)-butenedioic acid	<chem>C(O)(=O)C(C(Cl)Cl)=C(Cl)C(O)=O</chem>	Unknown	10Apr2006
55 DBPCAN_v3b_209_10Apr2006	(E)-2-Chloro-3-(dichloromethyl)-4-oxobutenoic acid	<chem>OC(C(Cl)=C(C=O)C(Cl)Cl)=O</chem>	Unknown	10Apr2006
58 DBPCAN_v3b_209_10Apr2006	1-Chloro-2-ethoxy-2-methoxy ethane	<chem>ClCC(OCC)OC</chem>	Unknown	10Apr2006

59 DBPCAN_v3b_209_10Apr2006	4-Chloro-3-keto-1-butanal	<chem>O=CCC(=O)CCl</chem>	Unknown	10Apr2006
61 DBPCAN_v3b_209_10Apr2006	2-Chloro-3-methyl-cis-butenedioic acid	<chem>OC(C(C)=C(C(O)=O)Cl)=O</chem>	Unknown	10Apr2006
67 DBPCAN_v3b_209_10Apr2006	1-Chloro-3,3,3-trichloro-1-propen-1-amine	<chem>ClC(Cl)(C=C(Cl)N)Cl</chem>	Unknown	10Apr2006
83 DBPCAN_v3b_209_10Apr2006	2,3-Dichloro-3-bromopropanenitrile	<chem>ClC(C(Cl)C#N)Br</chem>	Unknown	10Apr2006
85 DBPCAN_v3b_209_10Apr2006	3,4-Dichlorobutanenitrile	<chem>ClCC(C#N)Cl</chem>	Unknown	10Apr2006
92 DBPCAN_v3b_209_10Apr2006	4,5-Dichloro-2-pentanol	<chem>ClCC(Cl)CC(C)O</chem>	Unknown	10Apr2006
93 DBPCAN_v3b_209_10Apr2006	2,2-Dichloro-3-pentanone	<chem>CC(Cl)(Cl)C(=O)CC</chem>	Unknown	10Apr2006
98 DBPCAN_v3b_209_10Apr2006	Dihydro-4,5-dichloro-2-(3H)-furanone	<chem>O1C(Cl)C(Cl)CC1(=O)</chem>	Unknown	10Apr2006
107 DBPCAN_v3b_209_10Apr2006	1,2-Dioxopropanoic acid	<chem>OC(C(C=O)=O)=O</chem>	Unknown	10Apr2006
110 DBPCAN_v3b_209_10Apr2006	4-Dodecyl-5-ethyl-2(5H)furanone	<chem>O=C1C=C(CCCCCCCCCCCC)C(CC)O1</chem>	Unknown	10Apr2006
112 DBPCAN_v3b_209_10Apr2006	2-Ethyl-2-methyl-cis-butenedioic acid	<chem>O=C(O)C(C)=C(CC)C(O)=O</chem>	Unknown	10Apr2006
130 DBPCAN_v3b_209_10Apr2006	1-[4-(1-Hydroxy-1-methylethyl)phenyl]-ethanone	<chem>C1=CC(C(C)=O)=CC=C1(C(C)(O)C)</chem>	Unknown	10Apr2006
131 DBPCAN_v3b_209_10Apr2006	1-Hydroxy-3-methyl-2-hexene	<chem>C(CC)C(C)=CCO</chem>	Unknown	10Apr2006
132 DBPCAN_v3b_209_10Apr2006	5-Hydroxy-5-trichloromethyl-2-furanone	<chem>C1=CC(O)(C(Cl)(Cl)Cl)OC1(=O)</chem>	Unknown	10Apr2006
143 DBPCAN_v3b_209_10Apr2006	2-Methyl-3,3-dichloro-2-propenyl dichloromethyl ether	<chem>ClC(Cl)=C(COC(Cl)Cl)C</chem>	Unknown	10Apr2006
145 DBPCAN_v3b_209_10Apr2006	1-[4-(1-Methylethenyl) phenyl]-ethanone	<chem>C1=CC(C(C)=O)=CC=C1(C(C)=C)</chem>	Unknown	10Apr2006
189 DBPCAN_v3b_209_10Apr2006	1,1,1-Tribromo-2-bromo-2-chloroethane	<chem>BrC(Br)(Br)C(Cl)Br</chem>	Unknown	10Apr2006
192 DBPCAN_v3b_209_10Apr2006	1,1,1-Trichloro-2-butanone	<chem>O=C(CC)C(Cl)(Cl)Cl</chem>	Unknown	10Apr2006
193 DBPCAN_v3b_209_10Apr2006	cis-2,3,4-Trichloro-2-butenenitrile	<chem>N#CC(Cl)=C(Cl)CCl</chem>	Unknown	10Apr2006
194 DBPCAN_v3b_209_10Apr2006	trans-2,3,4-Trichloro-2-butenenitrile	<chem>N#CC(Cl)=C(Cl)CCl</chem>	Unknown	10Apr2006
197 DBPCAN_v3b_209_10Apr2006	5,5,5-Trichloro-4-oxopentanoic acid	<chem>O=C(C(Cl)(Cl)Cl)CCC(O)=O</chem>	Unknown	10Apr2006
206 DBPCAN_v3b_209_10Apr2006	1,3,3-Trimethyl-7-oxabicyclo-[4.1.0]-heptane-2,5-dione	<chem>C1(C)(C)CC(=O)C2OC2(C)C1(=O)</chem>	Unknown	10Apr2006